A Phylogenomic Analysis of the Origin of Plastids
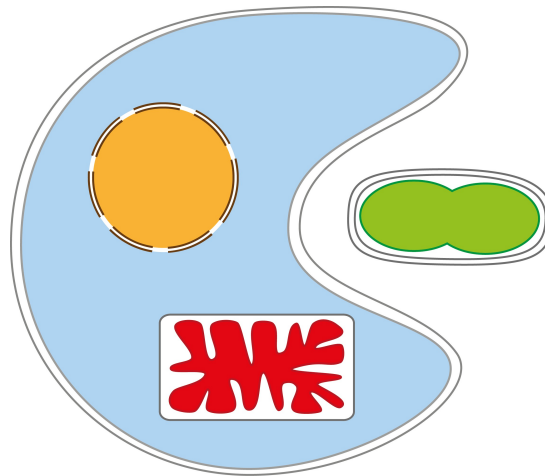
Luc Cornet [1][2], Emmanuelle Javaux [2], Annick Wilmotte [3], Hervé Philippe [4], Denis Baurain [1]

1 Eukaryotic Phylogenomics, University of Liège, Belgium
2 Palaeobiogeology-Palaeobotany-Palaeopalynology, University of Liège, Belgium
3 Centre for Protein Engineering, University of Liège, Belgium
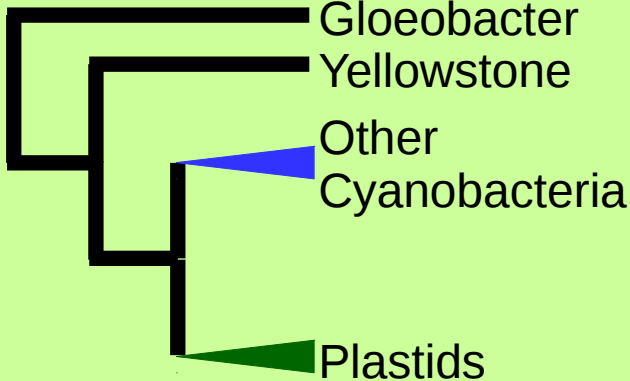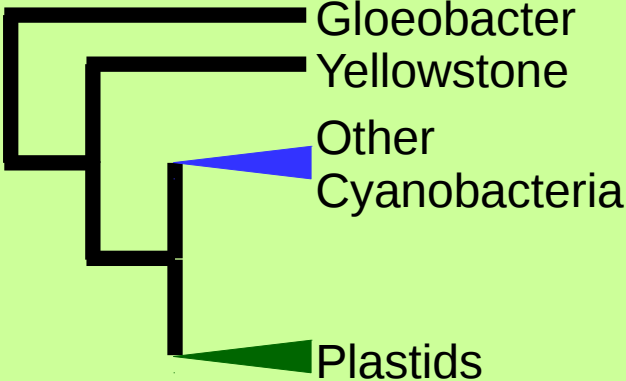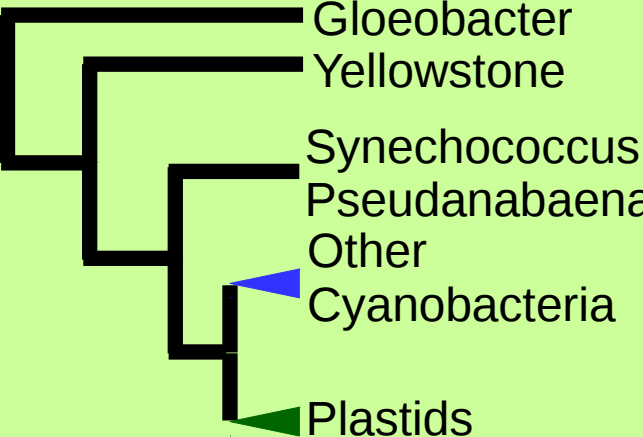4 Center for Biodiversity Theory and Modelling, USR CNRS 2936, France

Background
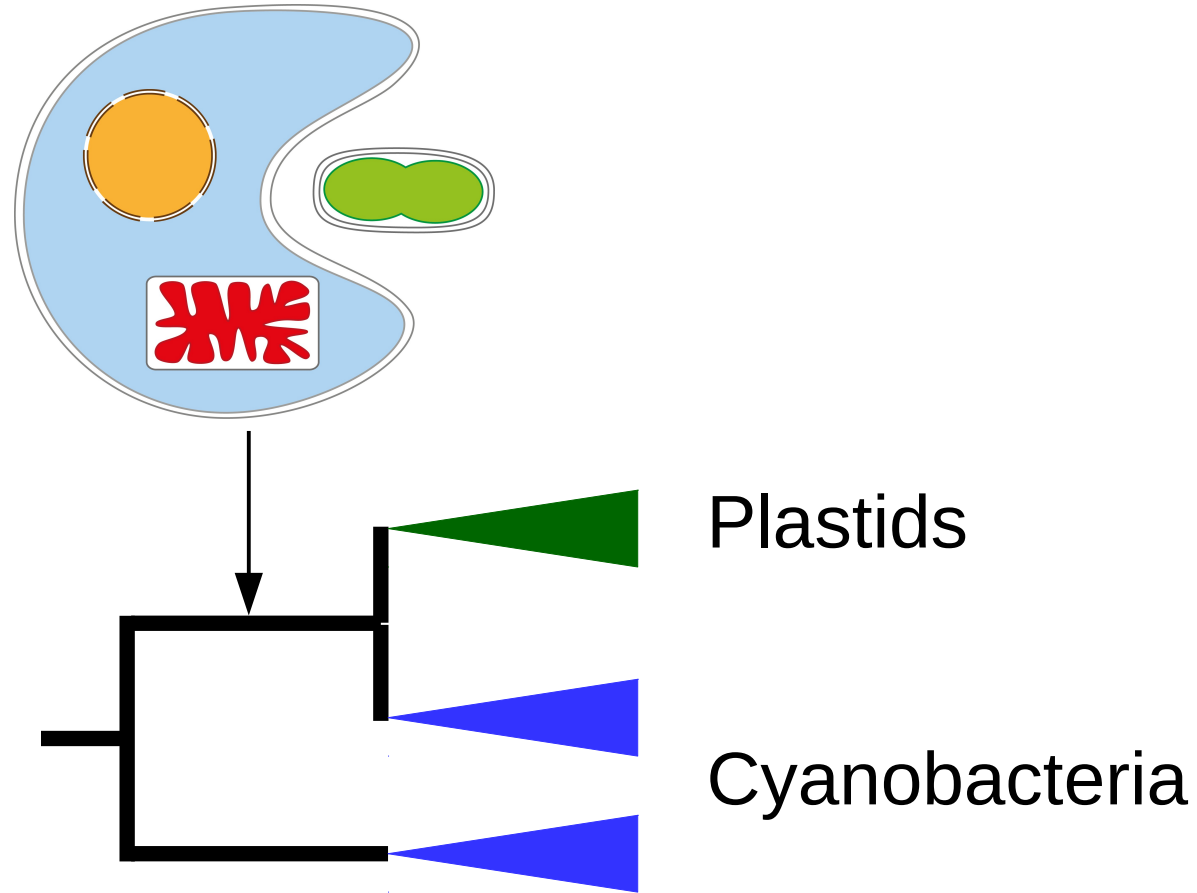
Plastids = Monophyly in Cyanobacteria

State of the Art

Criscuolo & Gribaldo., 2011	Li et al., 2014	Shih et al., 2013
Gloeobacter Yellowstone Other Cyanobacteria Plastids	Gloeobacter Yellowstone Other Cyanobacteria Plastids	Gloeobacter Yellowstone Synechococcus Pseudanabaena Other Cyanobacteria Plastids
61 Cyanobacteria 13 Archaeaplastida	16 Cyanobacteria 18 Archaeaplastida	126 Cyanobacteria 37 Archaeaplastida
Supermatrix GTR+G+I	Supermatrix CPREV+G+I	Supermatrix AA LG+G+I

Different positions of plastids

Objectives

- To determine the position of plastids using phylogenomic approaches

- Features of this work
  - Public genome data
  - Extensive taxon sampling (including close outgroups)
  - Sophisticated methods and evolutionary models
  - Good automation yet with careful manual controls
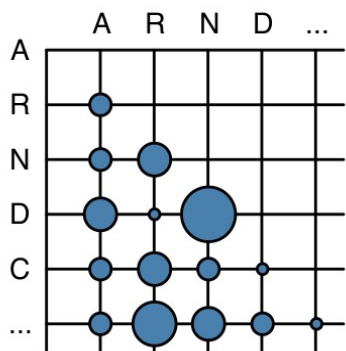
Materials

Ensembl   JGI   IMG   NCBI   GGK Base

105   20   14   1   101   8

19

Cyano
(140)

Plastids
(101)

Chloroflexi
(19)

Melainabacteria
(8)

Methods

all-*vs*-all Comparison → Using USEARCH (E-value 1e-5; minseqlength)

Gene clustering → Definition of orthologous groups (OGs)
for different values of inflation (1.1, 1.2, 1.3, 1.4, 1.5, 2)
using OrthoMCL pipeline

Annotation of plastid genes → Identification of OGs with plastid genes
by alignment against reference plastid genomes
using USEARCH global clustering

Determination of optimal clustering parameters → Selection of 313 single copy plastid-related OGs
containing at least 4 sequences

Alignment → Alignment of plastid-related OGs
using MAFFT

OTU Selection → Selection of 99 OTUs (present in at least 10 % OGs)
including a subset of 8 slowly evolving plastids
using SCaFoS

Gene concatenation → Selection and concatenation of 94 genes
with at most 15 missing OTUs
using Gblocks and SCaFoS

Phylogenetic inference → Analysis of multiple taxon-sampling variants
with different models (LG, CAT, CATGTR)
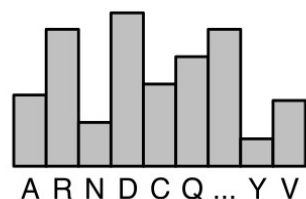using RAxML and PhyloBayes
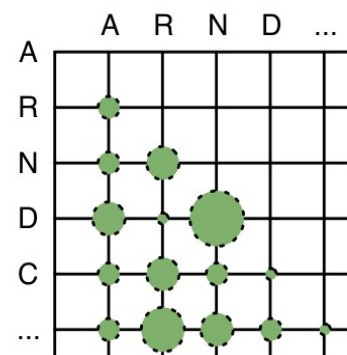
ASSEMBLY

Evolutionary Models

LG or WAG

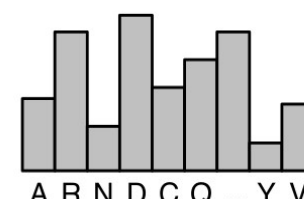1 global precomputed replacement matrix + 1 global compositional profile
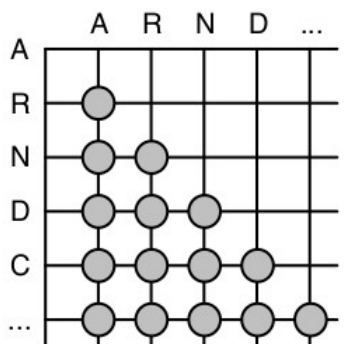
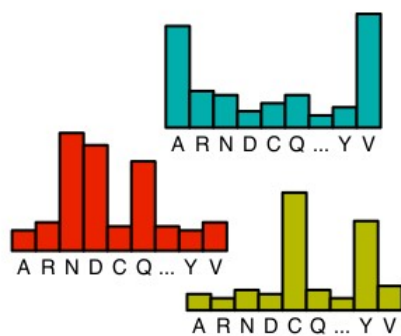GTR

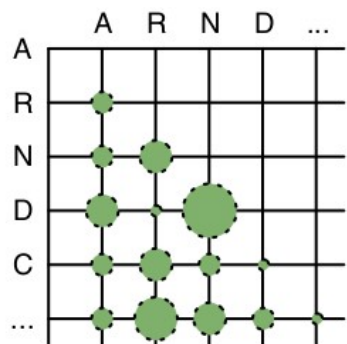1 global « dynamic » replacement matrix + 1 global compositional profile
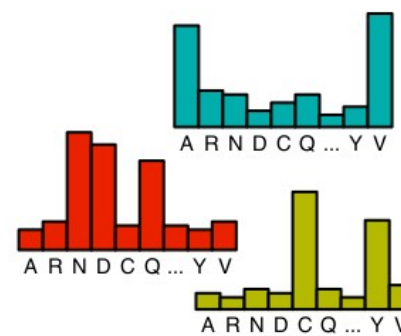
CAT

1 global « flat » replacement matrix + *K* distinct compositional profiles

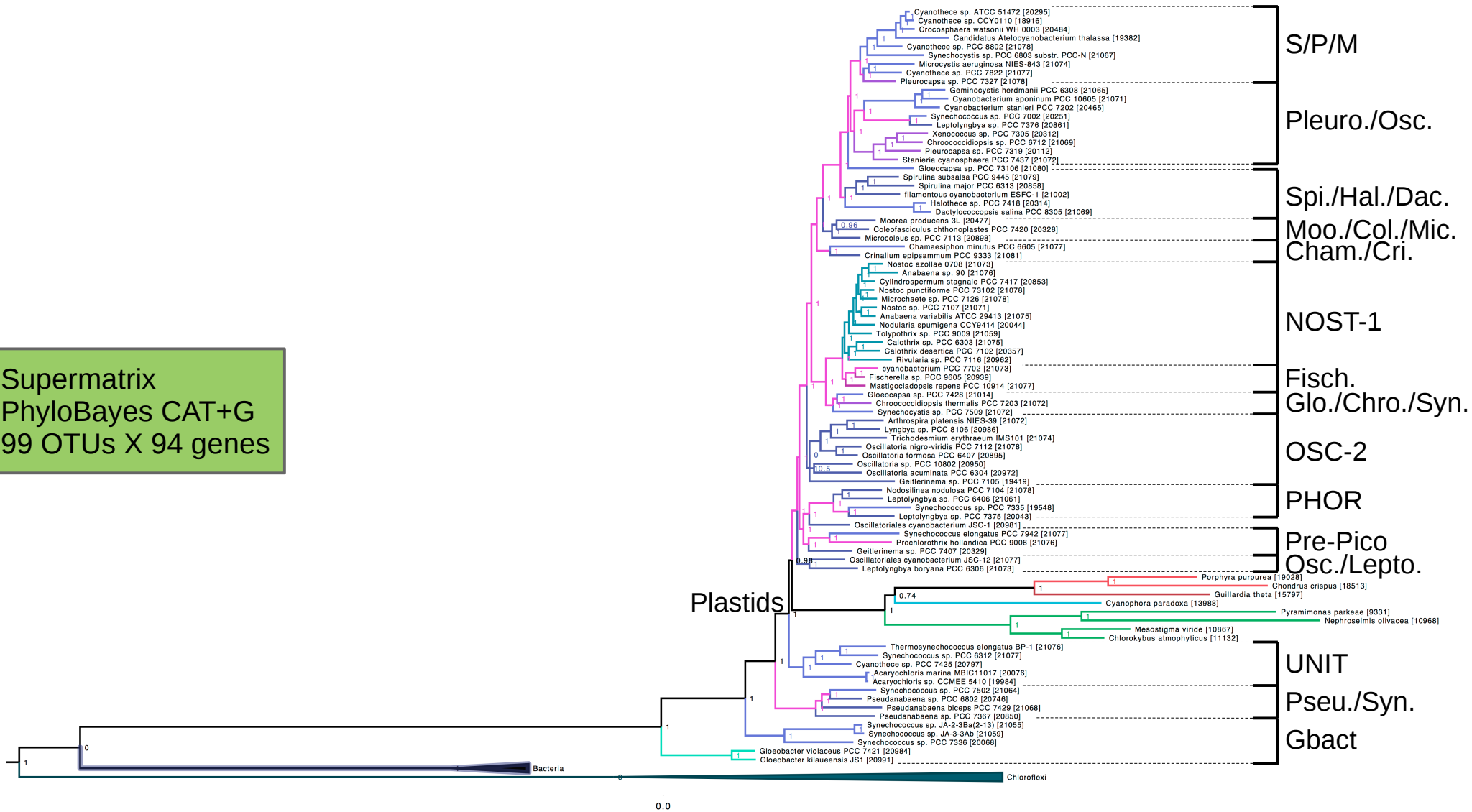CATGTR

1 global « dynamic » replacement matrix + *K* distinct compositional profiles

Tree: **Plastid Supermatrix**

Supermatrix
PhyloBayes CAT+G
99 OTUs X 94 genes

Similar topology in LG, CAT and CATGTR

Results
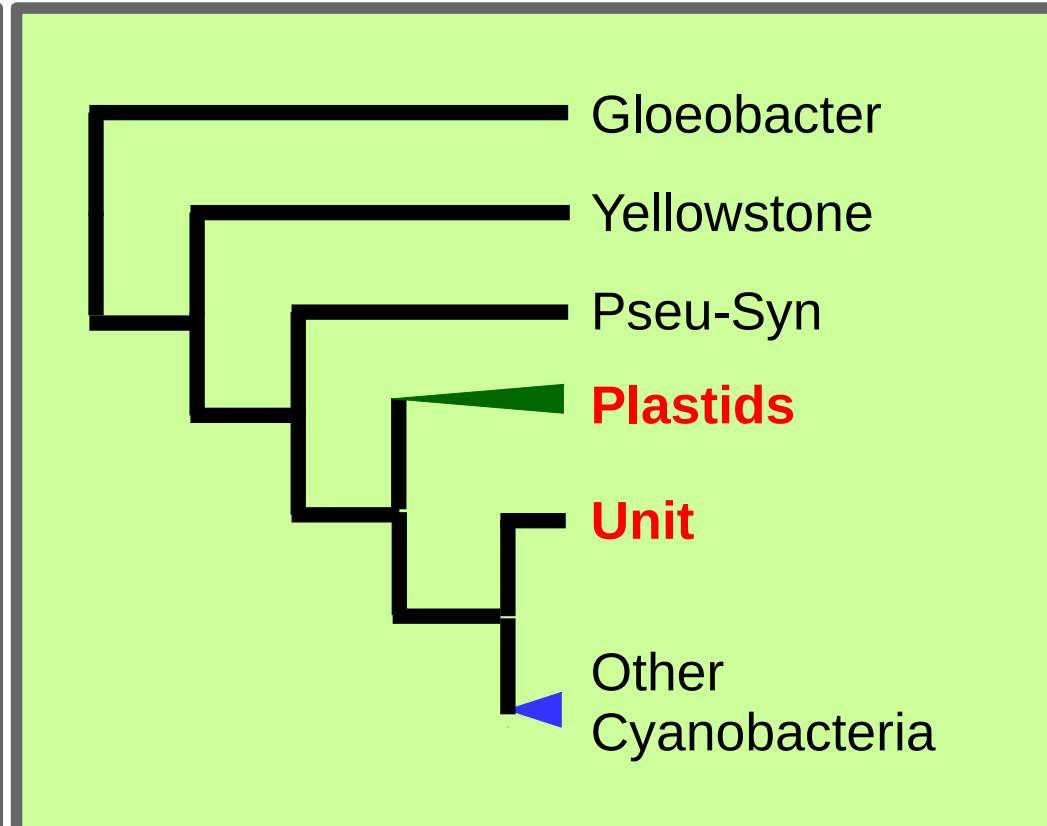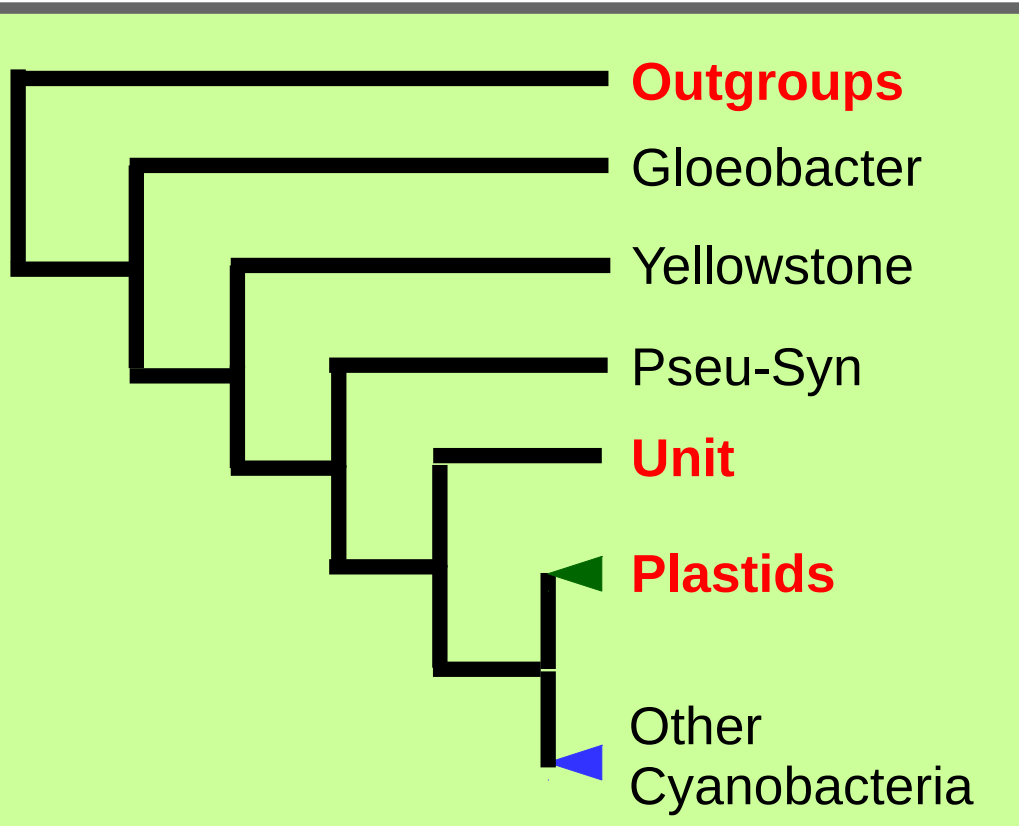
Unstable position of plastids across taxon sampling variants: phylogenetic artefact?
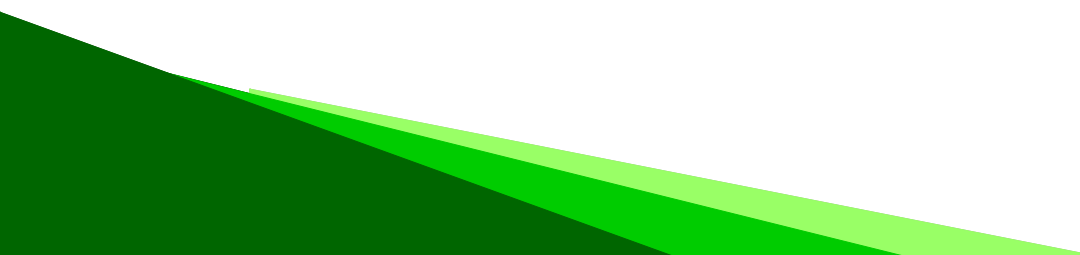
Intermediate Conclusions

- Not so early origin of plastids
  - Ongoing: analysis of phylogenetic artefacts (compositional/saturational tests and tests for heterotachy/heteropecilly using posterior prediction in PhyloBayes)
  - To do: removal of fast evolving sites; analysis of gene sampling variants (jackknife)

- Computational considerations
  CAT = 1 month of CPU time
  CATGTR = 32 months of CPU time

Need for corroboration

- Change methods

- Change datasets

Change Methods: Supertrees

1. Matrix representation with parsimony (MRP)

2. Average Consensus (Av cons)

3. Subtree prune-and-regraft (SPR) distance

Change Methods: Supertrees

Supertree LG+F+G
Average Consensus
94 OTUs x 94 genes

Plastids

Plastid position as with supermatrix

Change Methods: Supertrees

Supertree LG+F+G
Average Consensus
94 OTUs x 94 genes

Outgroup
GBACT
Pseud-Syn
Plastids
Unit
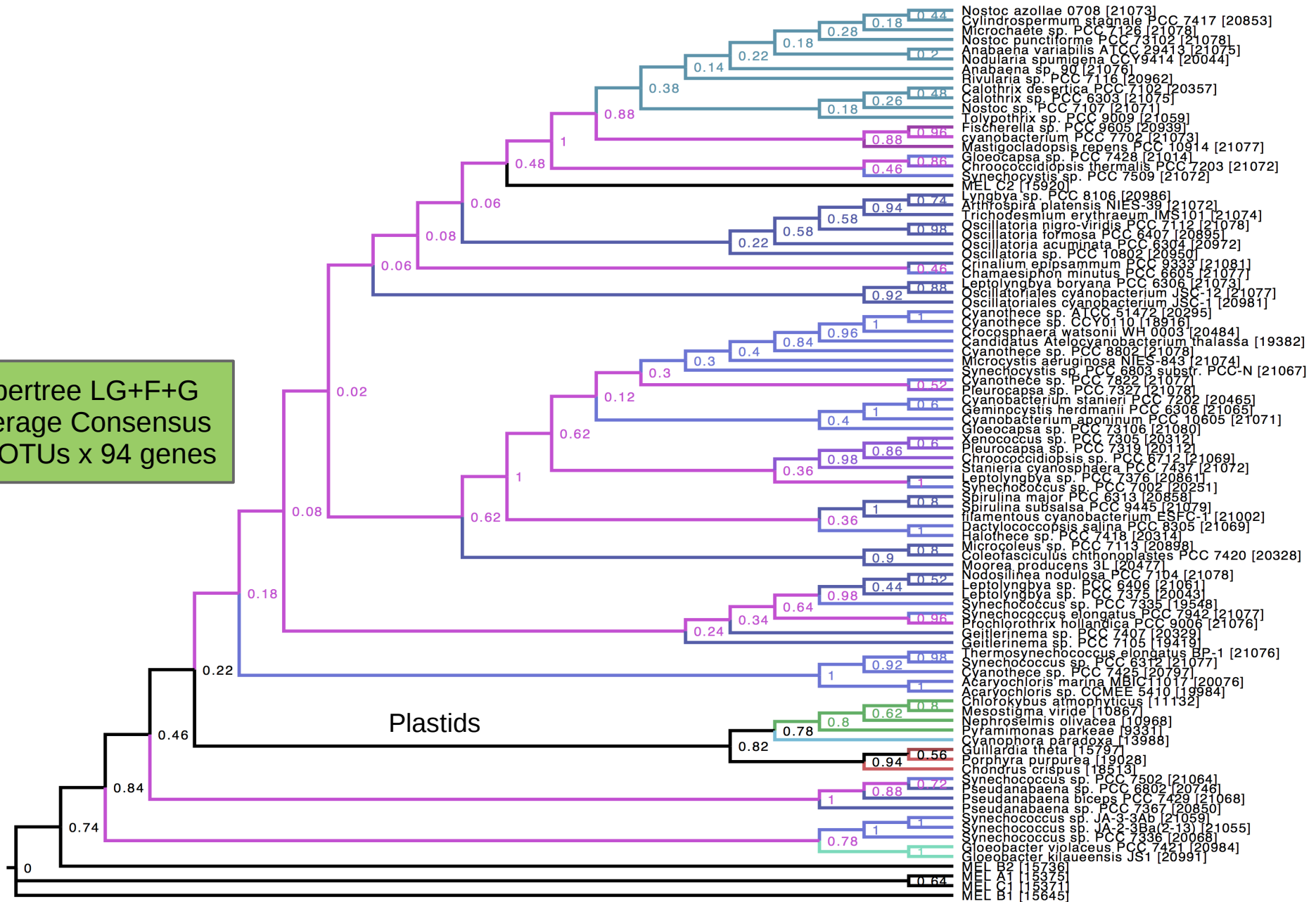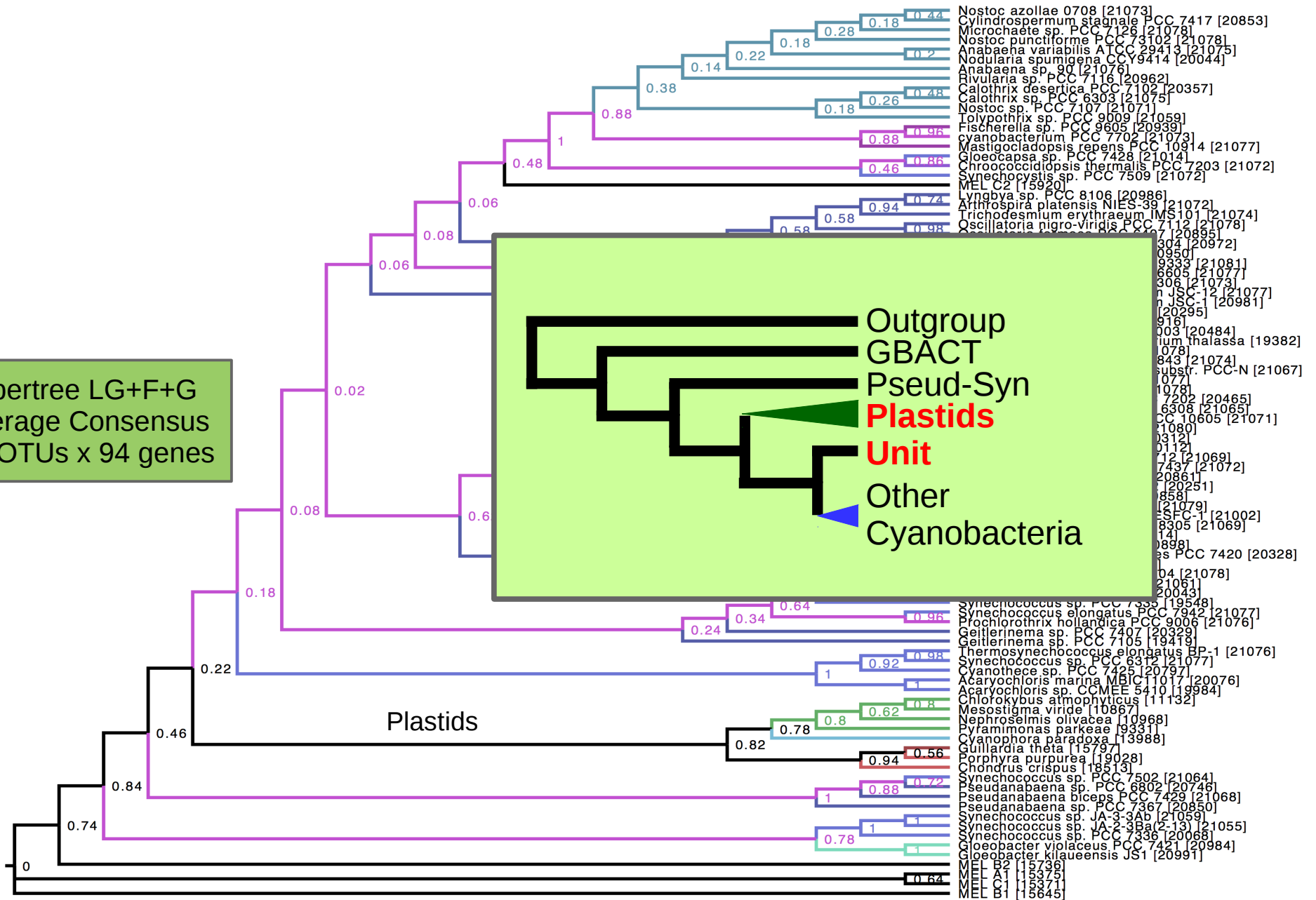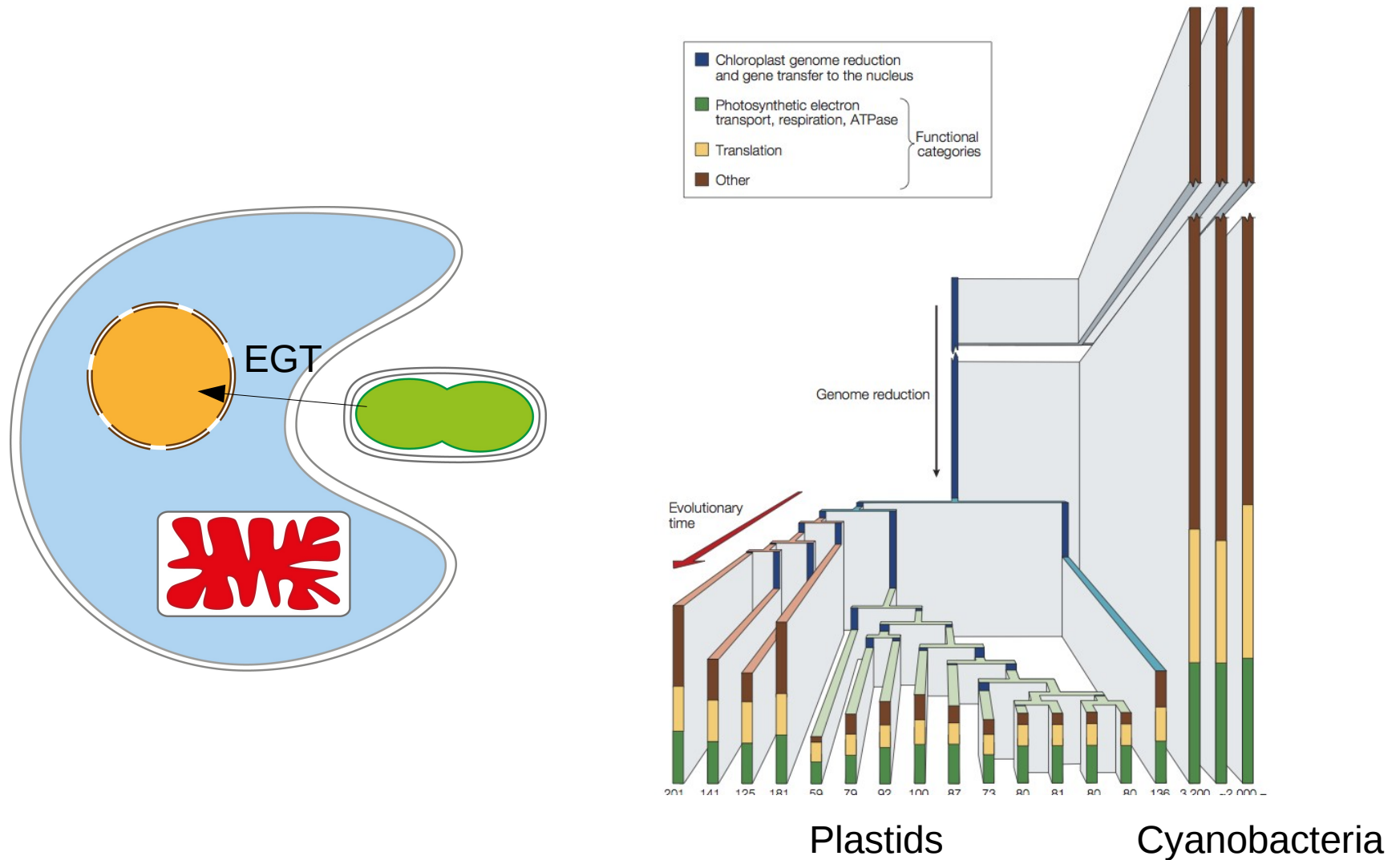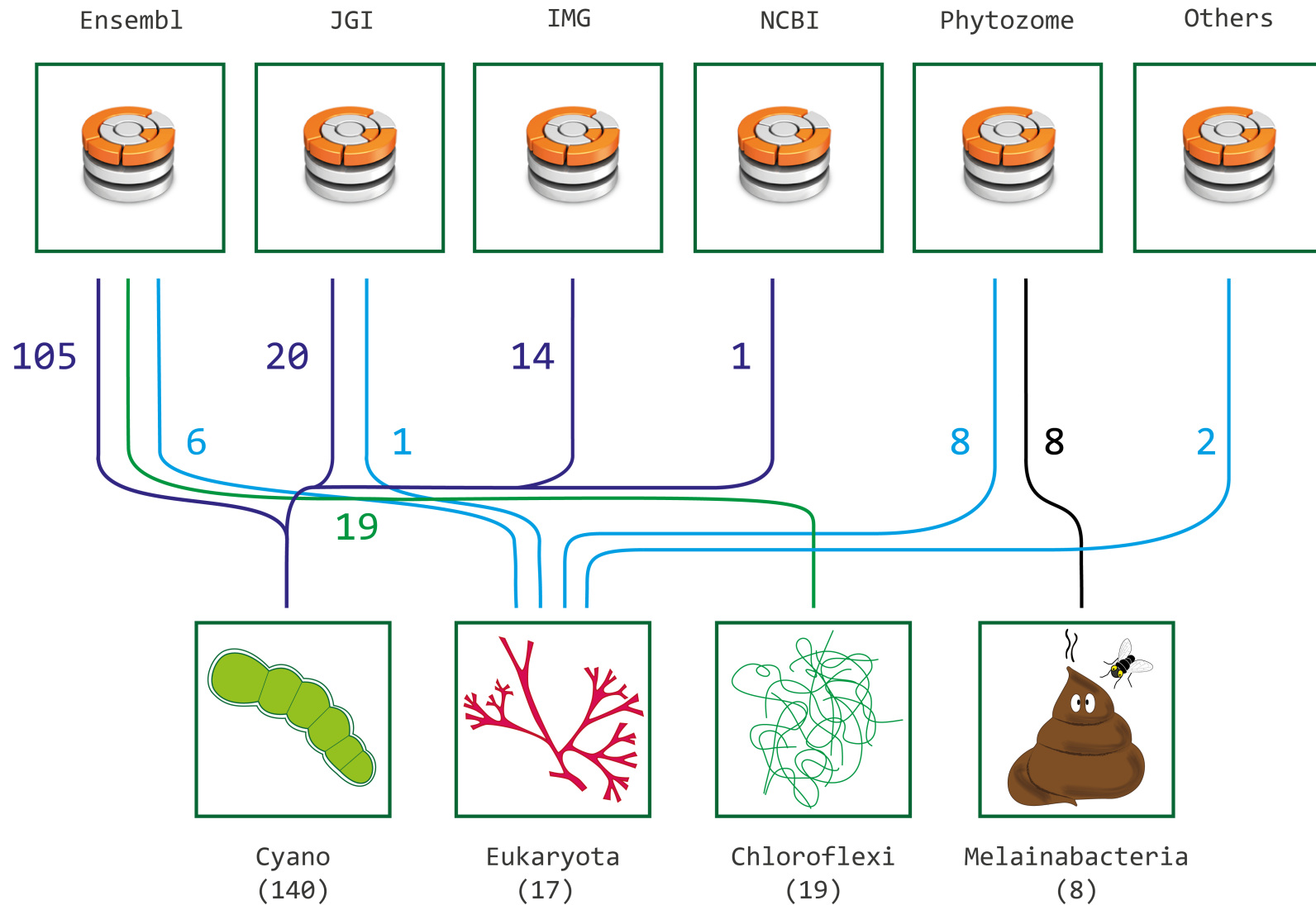Other
Cyanobacteria

Plastids

Plastid position as with supermatrix

Nostoc azollae 0708 [21073]
Cylindrospermum stagnale PCC 7417 [20853]
Microchaete sp. PCC 7126 [21078]
Nostoc punctiforme PCC 73102 [21078]
Anabaena variabilis ATCC 29413 [21075]
Nodularia spumigena CCY9414 [20044]
Anabaena sp. 90 [21076]
Rivularia sp. PCC 7116 [20962]
Calothrix desertica PCC 7102 [20357]
Calothrix sp. PCC 6303 [21075]
Nostoc sp. PCC 7107 [21071]
Tolypothrix sp. PCC 9009 [21059]
Fischerella sp. PCC 9605 [20939]
cyanobacterium PCC 7702 [21073]
Mastigocladopsis repens PCC 10914 [21077]
Gloeocapsa sp. PCC 7428 [21014]
Chroococcidiopsis thermalis PCC 7203 [21072]
Synechocystis sp. PCC 7509 [21072]
MEL C2 [15920]
Lyngbya sp. PCC 8106 [20986]
Arthrospira platensis NIES-39 [21072]
Trichodesmium erythraeum IMS101 [21074]
Oscillatoria nigro-viridis PCC 7112 [21078]

Synechococcus sp. PCC 7335 [19548]
Synechococcus elongatus PCC 7942 [21077]
Prochlorothrix hollandica PCC 9006 [21076]
Geitlerinema sp. PCC 7407 [20329]
Geitlerinema sp. PCC 7105 [19419]
Thermosynechococcus elongatus BP-1 [21076]
Synechococcus sp. PCC 6312 [21077]
Cyanothece sp. PCC 7425 [20797]
Acaryochloris marina MBIC11017 [20076]
Acaryochloris sp. CCMEE 5410 [19984]
Chlorokybus atmophyticus [11132]
Mesostigma viride [10867]
Nephroselmis olivacea [10968]
Pyramimonas parkeae [9331]
Cyanophora paradoxa [13988]
Guillardia theta [15797]
Porphyra purpurea [19028]
Chondrus crispus [18513]
Synechococcus sp. PCC 7502 [21064]
Pseudanabaena sp. PCC 6802 [20746]
Pseudanabaena biceps PCC 7429 [21068]
Pseudanabaena sp. PCC 7367 [20850]
Synechococcus sp. JA-3-3Ab [21059]
Synechococcus sp. JA-2-3Ba(2-13) [21055]
Synechococcus sp. PCC 7336 [20068]
Gloeobacter violaceus PCC 7421 [20984]
Gloeobacter kilaueensis JS1 [20991]
MEL B2 [15736]
MEL A1 [15375]
MEL C1 [15371]
MEL B1 [15645]

Change Dataset: Nuclear Genes
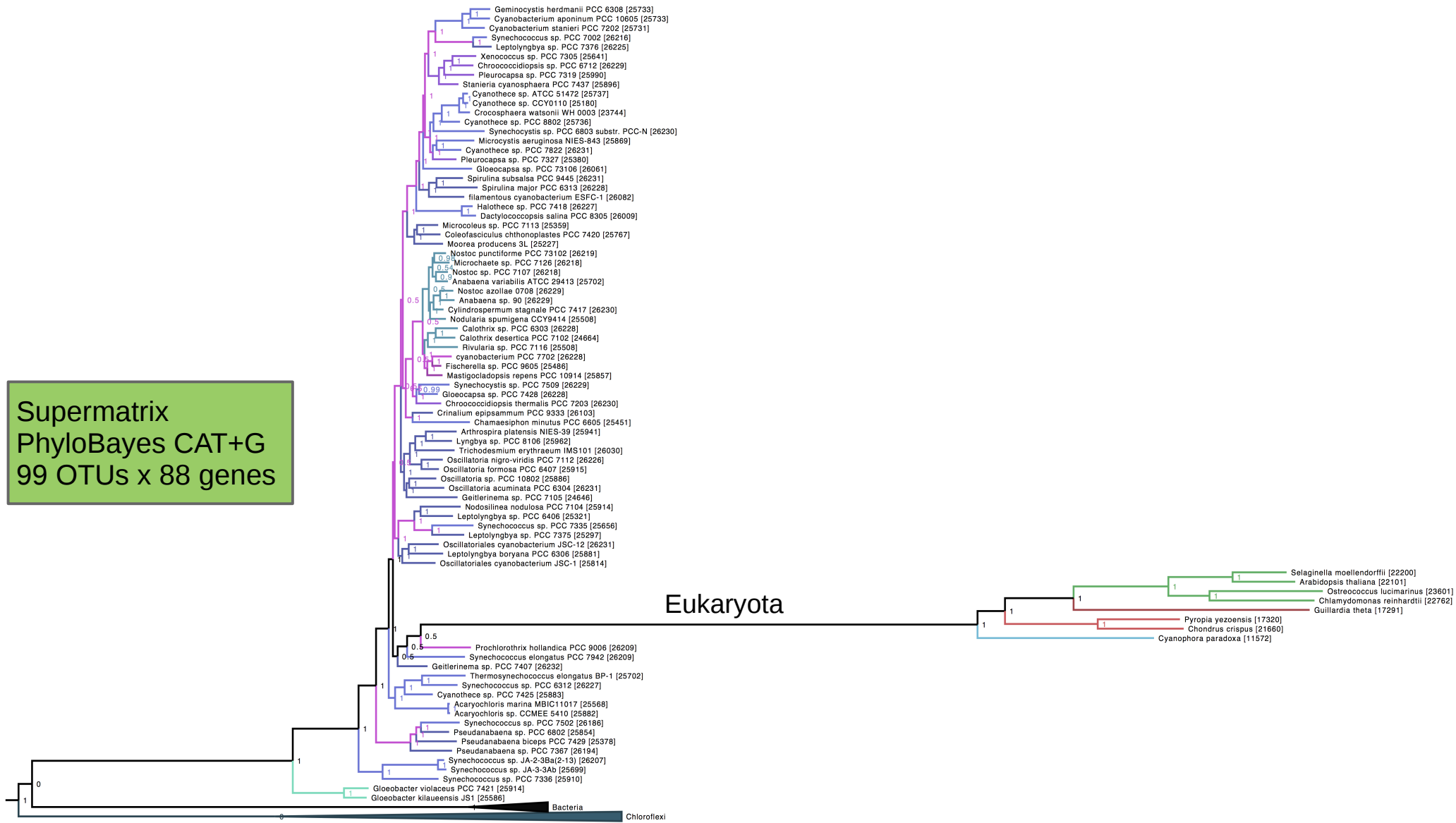

Nuclear genes of endosymbiotic origin

Change Dataset: Nuclear Genes

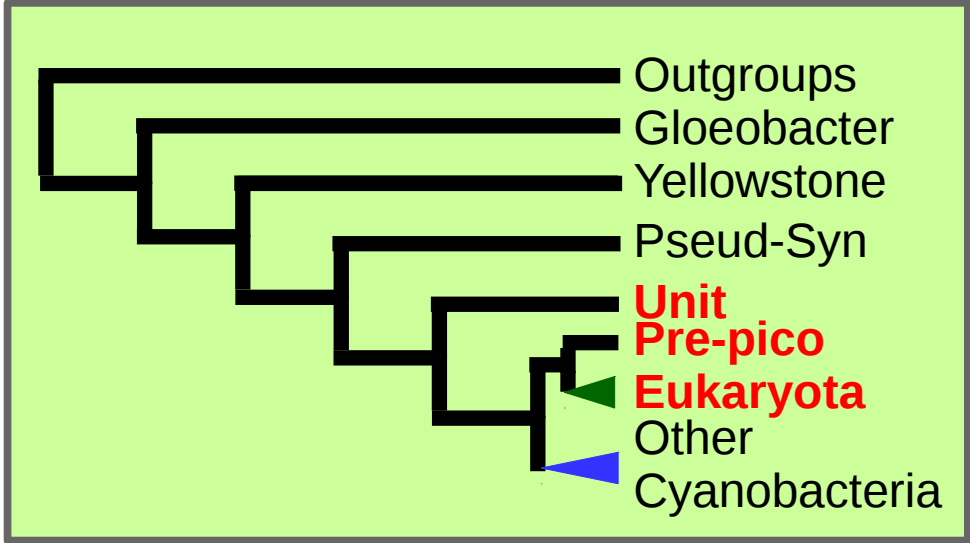Same pipeline as for plastid dataset
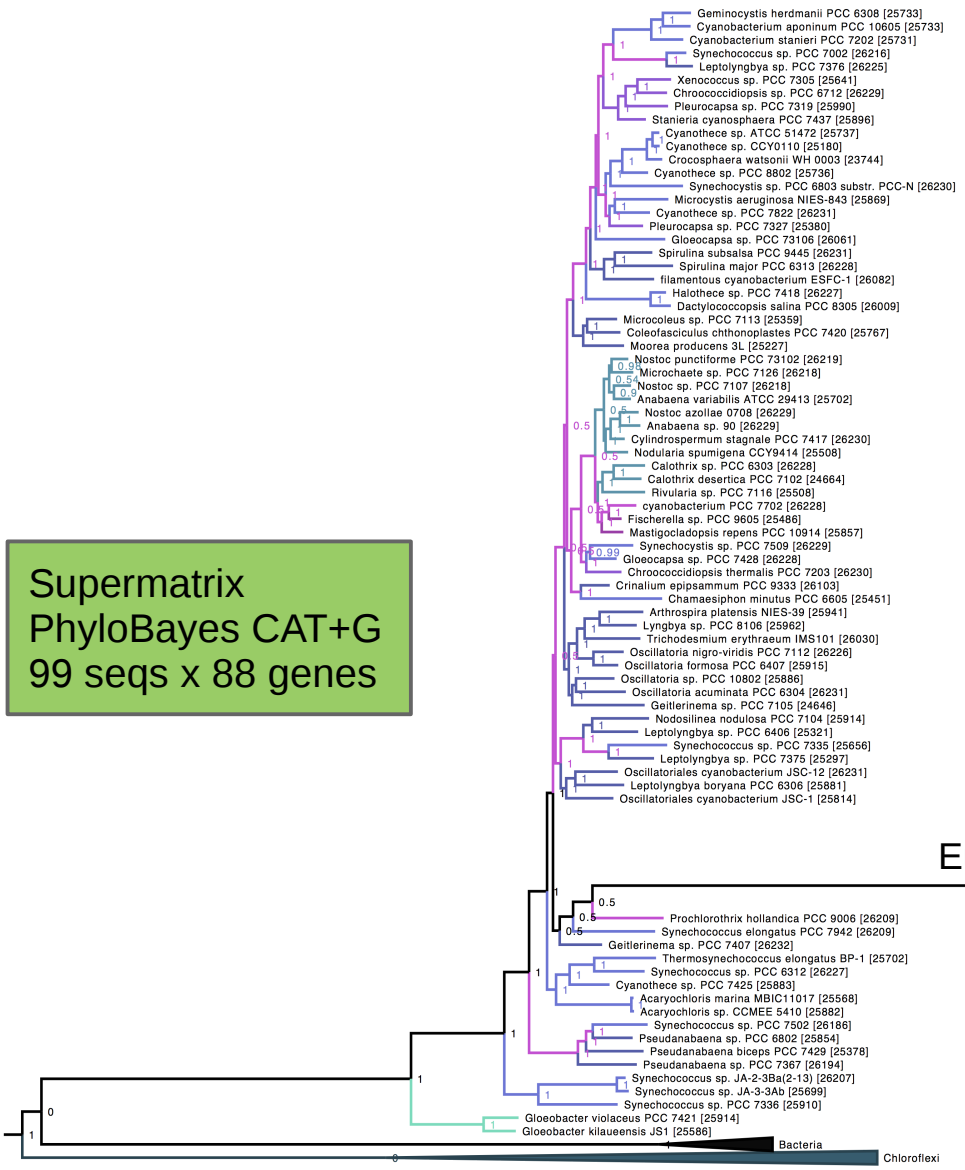
Change Dataset: Nuclear Genes
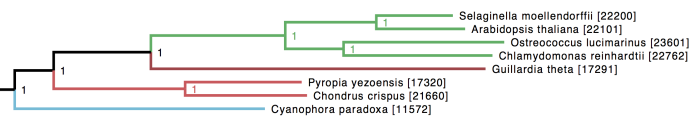

Supermatrix
PhyloBayes CAT+G
99 OTUs x 88 genes

Eukaryota

Plastids position similar to plastid dataset

Change Dataset: Nuclear Genes

Supermatrix
PhyloBayes CAT+G
99 seqs x 88 genes

Plastids position similar to plastid dataset

General Conclusions

- Use of two different datasets corresponding to two gene classes (plastid- and nuclear-encoded)
- Use of two different phylogenomic approaches

- Not so early origin of plastids but still to be demonstrated

Perspectives

Sequencing of private Antarctic strains (broadly sampled), focus on the candidate sister groups (Gbact, Pseu./Syn.,Unit, Osc./Lepto.)

Thank you for your attention